Artificial Intelligence Nanodegree
Assignment 2 – Advanced Game Playing – Research Review
Charles Ocheret
February 19, 2017

AlphaGo – Summary of "Mastering the Game of Go with Deep Neural Networks and Tree Search"

The Goals of the Paper

The authors of the paper present implementation details and observations of the first automated player of the game Go capable of defeating expert human players. Go is as 2,500 year old Chinese game originally from China. Despite having relatively simple rules, the complexity of the game explodes quickly as the breadth and depth of the search space for Go is roughly 250^{150} (as compared to Chess, which is roughly 35^{80}). The authors beat the community's estimates for a competitive AI to play Go at this level by 10 years. Interestingly, it appears the authors primary breakthrough is the extremely clever application of well understood techniques in a novel way, in particular Convolutional Neural Networks and Monte Carlo Tree Search.

Summary of the Techniques Employed

Convolutional Neural Networks (CNNs) were originally modeled after the animal visual cortex system and has been successfully applied to problems like image recognition. CNNs group a grid of inputs into a grid of overlapping tiles that can each be trained to recognize patterns (effectively implemented as a 2D convolution kernel for each pattern). This offers properties such as translational independence (i.e. the same pattern can be recognized no matter which tile it appears in). The pattern of connectivity can is repeated for many levels (using rectifiers between levels) to offer recognition of increasingly complex patterns. Applying CNNs to the Go playing board appears to be a novel idea. Yet it makes sense since the 19x19 board is a small grid (compared to large images) and recognizable patterns of different scales on a Go board are natural features to look for.

Monte Carlo Tree Search is a tree search algorithm that explores a subset of a large tree based on random selection of paths to leaf nodes to build up an increasingly accurate measure of the value of nodes for the improved performance of future searches.

CNNs are used in two distinct ways. A value network is used to reduce search depth by providing good estimates of the value of each node. A policy network reduces the search breadth by selecting the best action for each node based on a probability distribution function. The policy network is bootstrapped using a supervised learning (SL) based on the gameplay of humans. A lower quality policy network is also derived for much faster tree traversals required

by a MCTS that is used to train the value network. A parallel policy network is then bootstrapped from the initial SL network and is refined through additional computer gameplay. This is the Reinforcement Learning (RL) policy network. During gameplay the value network is used in combination with MCTS rollouts to refine action values.

Results

AlphaGo was able to win matches against expert players such as Fan Hui and Lee Sedol, which is a tremendous achievement in itself and a milestone in the advancement of the state of the art in Artificial Intelligence. The implementation has been adapted to run across CPUs and GPUs on powerful individual machines and on a distributed network of machines. The clever combination of technologies allows AlphaGo to make effective decisions after far less searching of the search space than other approaches (e.g. DeepBlue's chess algorithm). The expectation is that these techniques can be brought to bear on many other problems that had been deemed intractable before.